

WILEY-VCH

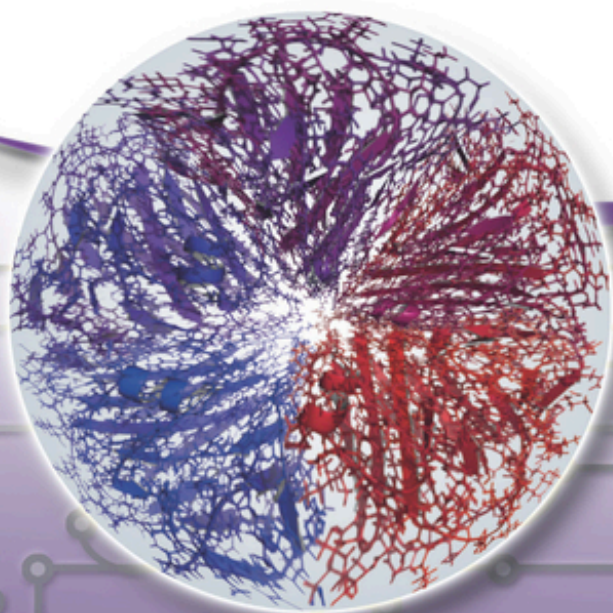
Edited by
Huimin Zhao

Protein Engineering

Tools and Applications

Volume 10

Series Editors:
S. Y. Lee, J. Nielsen,
G. Stephanopoulos



Advanced Biotechnology

Protein Engineering

Protein Engineering

Tools and Applications

Edited by
Huimin Zhao

WILEY-VCH

Volume Editor**Professor Huimin Zhao**

University of Illinois at Urbana
Chemical & Biomolecular Engineering
600 South Mathews Avenue
215 Roger Adams Laboratory
61801 Urbana IL
USA

Series Editors**Prof. Dr. Sang Yup Lee**

KAIST
373-1; Guseong-Dong
291 Daehak-ro, Yuseong-gu
305-701 Daejeon
South Korea

Prof. Dr. Jens Nielsen

Chalmers University
Department of Biology and Biological
Engineering
Kemivägen 10
412 96 Göteborg
Sweden

Prof. Dr. Gregory Stephanopoulos

Massachusetts Institute of Technology
Department of Chemical Engineering
Massachusetts Ave 77
Cambridge, MA 02139
USA

Cover Culture Flasks in microbiological
laboratory / science photo, fotolia

■ All books published by **WILEY-VCH** are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.:
applied for

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <<http://dnb.d-nb.de>>.

© 2021 WILEY-VCH GmbH, Boschstr. 12, 69469 Weinheim, Germany

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Print ISBN: 978-3-527-34470-3

ePDF ISBN: 978-3-527-81509-8

ePub ISBN: 978-3-527-81511-1

oBook ISBN: 978-3-527-81512-8

Cover Design Adam-Design, Weinheim, Germany

Typesetting Straive, Chennai, India

Printing and Binding

Printed on acid-free paper

10 9 8 7 6 5 4 3 2 1

Contents

Part I Directed Evolution 1

1	Continuous Evolution of Proteins <i>In Vivo</i>	3
	<i>Alon Wellner, Arjun Ravikumar, and Chang C. Liu</i>	
1.1	Introduction	3
1.2	Challenges in Achieving <i>In Vivo</i> Continuous Evolution	5
1.3	Phage-Assisted Continuous Evolution (PACE)	10
1.4	Systems That Allow <i>In Vivo</i> Continuous Directed Evolution	13
1.4.1	Targeted Mutagenesis in <i>E. coli</i> with Error-Prone DNA Polymerase I	13
1.4.2	Yeast Systems That Do Not Use Engineered DNA Polymerases for Mutagenesis	16
1.4.3	Somatic Hypermutation as a Means for Targeted Mutagenesis of GOIs	18
1.4.4	Orthogonal DNA Replication (OrthoRep)	20
1.5	Conclusion	22
	References	22
2	<i>In Vivo</i> Biosensors for Directed Protein Evolution	29
	<i>Song Buck Tay and Ee Lui Ang</i>	
2.1	Introduction	29
2.2	Nucleic Acid-Based <i>In Vivo</i> Biosensors for Directed Protein Evolution	32
2.2.1	RNA-Type Biosensors	32
2.2.2	DNA-Type Biosensors	35
2.3	Protein-Based <i>In Vivo</i> Biosensors for Directed Protein Evolution	37
2.3.1	Transcription Factor-Type Biosensors	37
2.3.2	Enzyme-Type Biosensors	41
2.4	Characteristics of Biosensors for <i>In Vivo</i> Directed Protein Evolution	44
2.5	Conclusions and Future Perspectives	45
	Acknowledgments	46
	References	46

3	High-Throughput Mass Spectrometry Complements Protein Engineering 57
	<i>Tong Si, Pu Xue, Kisurb Choe, Huimin Zhao, and Jonathan V. Sweedler</i>
3.1	Introduction 57
3.2	Procedures and Instrumentation for MS-Based Protein Assays 59
3.3	Technology Advances Focusing on Throughput Improvement 62
3.4	Applications of MS-Based Protein Assays: Summary 63
3.4.1	Applications of MS-Based Assays: Protein Analysis 64
3.4.2	Applications of MS-Based Assays: Protein Engineering 66
3.5	Conclusions and Perspectives 68
	Acknowledgments 68
	References 69
4	Recent Advances in Cell Surface Display Technologies for Directed Protein Evolution 81
	<i>Maryam Raeeszadeh-Sarmazdeh and Wilfred Chen</i>
4.1	Cell Display Methods 81
4.1.1	Phage Display 81
4.1.2	Bacterial Display Systems 83
4.1.3	Yeast Surface Display 84
4.1.4	Mammalian Display 85
4.2	Selection Methods and Strategies 86
4.2.1	High-Throughput Cell Screening 86
4.2.1.1	Panning 86
4.2.1.2	FACS 86
4.2.1.3	MACS 87
4.2.2	Selection Strategies 88
4.2.2.1	Competitive Selection (Counter Selection) 88
4.2.2.2	Negative/Positive Selection 89
4.3	Modifications of Cell Surface Display Systems 89
4.3.1	Modification of YSD for Enzyme Engineering 89
4.3.2	Yeast Co-display System 91
4.3.3	Surface Display of Multiple Proteins 91
4.4	Recent Advances to Expand Cell-Display Directed Evolution Techniques 93
4.4.1	μ SCALE (Microcapillary Single-Cell Analysis and Laser Extraction) 93
4.4.2	Combining Cell Surface Display and Next-Generation Sequencing 94
4.4.3	PACE (Phage-Assisted Continuous Evolution) 94
4.5	Conclusion and Outlook 96
	References 97
5	Iterative Saturation Mutagenesis for Semi-rational Enzyme Design 105
	<i>Ge Qu, Zhoutong Sun, and Manfred T. Reetz</i>
5.1	Introduction 105
5.2	Recent Methodology Developments in ISM-Based Directed Evolution 108

- 5.2.1 Choosing Reduced Amino Acid Alphabets Properly 109
- 5.2.1.1 Limonene Epoxide Hydrolase as the Catalyst in Hydrolytic Desymmetrization 109
- 5.2.1.2 Alcohol Dehydrogenase TbSADH as the Catalyst in Asymmetric Transformation of Difficult-to-Reduce Ketones 110
- 5.2.1.3 P450-BM3 as the Chemo- and Stereoselective Catalyst in a Whole-Cell Cascade Sequence 112
- 5.2.1.4 Multi-parameter Evolution Aided by Mutability Landscaping 115
- 5.2.2 Further Methodology Developments of CAST/ISM 117
- 5.2.2.1 Advances Based on Novel Molecular Biological Techniques and Computational Methods 117
- 5.2.2.2 Advances Based on Solid-Phase Chemical Synthesis of SM Libraries 118
- 5.3 B-FIT as an ISM Method for Enhancing Protein Thermostability 120
- 5.4 Learning from CAST/ISM-Based Directed Evolution 121
- 5.5 Conclusions and Perspectives 121
- Acknowledgment 124
- References 124

Part II Rational and Semi-Rational Design 133

- 6 Data-driven Protein Engineering 135**
Jonathan Greenhalgh, Apoorv Saraogee, and Philip A. Romero
- 6.1 Introduction 135
- 6.2 The Data Revolution in Biology 136
- 6.3 Statistical Representations of Protein Sequence, Structure, and Function 138
- 6.3.1 Representing Protein Sequences 138
- 6.3.2 Representing Protein Structures 140
- 6.4 Learning the Sequence-Function Mapping from Data 141
- 6.4.1 Supervised Learning (Regression/Classification) 141
- 6.4.2 Unsupervised/Semisupervised Learning 144
- 6.5 Applying Statistical Models to Engineer Proteins 145
- 6.6 Conclusions and Future Outlook 147
- References 148
- 7 Protein Engineering by Efficient Sequence Space Exploration Through Combination of Directed Evolution and Computational Design Methodologies 153**
Subrata Pramanik, Francisca Contreras, Mehdi D. Davari, and Ulrich Schwaneberg
- 7.1 Introduction 153
- 7.2 Protein Engineering Strategies 154
- 7.2.1 Computer-Aided Rational Design 155
- 7.2.1.1 FRESCO 155
- 7.2.1.2 FoldX 157

- 7.2.1.3 CNA 158
- 7.2.1.4 PROSS 159
- 7.2.1.5 ProSAR 160
- 7.2.2 Knowledge Based Directed Evolution 161
- 7.2.2.1 Iterative Saturation Mutagenesis (ISM) 161
- 7.2.2.2 Mutagenic Organized Recombination Process by Homologous *In Vivo* Grouping (MORPHING) 161
- 7.2.2.3 Knowledge Gaining Directed Evolution (KnowVolution) 162
- 7.3 Conclusions and Future Perspectives 171
- Acknowledgments 171
- References 171

8 Engineering Artificial Metalloenzymes 177

Kevin A. Harnden, Yajie Wang, Lam Vo, Huimin Zhao, and Yi Lu

- 8.1 Introduction 177
- 8.2 Rational Design 177
- 8.2.1 Rational Design of Metalloenzymes Using *De Novo* Designed Scaffolds 177
- 8.2.2 Rational Design of Metalloenzymes Using Native Scaffolds 179
- 8.2.2.1 Redesign of Native Proteins 179
- 8.2.2.2 Cofactor Replacement in Native Proteins 181
- 8.2.2.3 Covalent Anchoring in Native Protein 184
- 8.2.2.4 Supramolecular Anchoring in Native Protein 187
- 8.3 Engineering Artificial Metalloenzyme by Directed Evolution in Combination with Rational Design 188
- 8.3.1 Directed Evolution of Metalloenzymes Using *De Novo* Designed Scaffolds 188
- 8.3.2 Directed Evolution of Metalloenzymes Using Native Scaffolds 189
- 8.3.2.1 Cofactor Replacement in Native Proteins 189
- 8.3.2.2 Covalent Anchoring in Native Protein 192
- 8.3.2.3 Non-covalent Anchoring in Native Proteins 194
- 8.4 Summary and Outlook 200
- Acknowledgment 201
- References 201

9 Engineered Cytochromes P450 for Biocatalysis 207

Hanan Alwaseem and Rudi Fasan

- 9.1 Cytochrome P450 Monooxygenases 207
- 9.2 Engineered Bacterial P450s for Biocatalytic Applications 210
- 9.2.1 Oxyfunctionalization of Small Organic Substrates 211
- 9.2.2 Late-Stage Functionalization of Natural Products 220
- 9.2.3 Synthesis of Drug Metabolites 224
- 9.3 High-throughput Methods for Screening Engineered P450s 227
- 9.4 Engineering of Hybrid P450 Systems 229
- 9.5 Engineered P450s with Improved Thermostability and Solubility 230

- 9.6 Conclusions 231
- Acknowledgments 232
- References 232

Part III Applications in Industrial Biotechnology 243

10 Protein Engineering Using Unnatural Amino Acids 245

Yang Yu, Xiaohong Liu, and Jiangyun Wang

- 10.1 Introduction 245
- 10.2 Methods for Unnatural Amino Acid Incorporation 246
- 10.3 Applications of Unnatural Amino Acids in Protein Engineering 247
 - 10.3.1 Enhancing Stability 248
 - 10.3.2 Mechanistic Study Using Spectroscopic Methods 248
 - 10.3.3 Tuning Catalytic Activity 250
 - 10.3.4 Tuning Selectivity 252
 - 10.3.5 Enzyme Design 252
 - 10.3.6 Protein Engineering Toward a Synthetic Life 255
- 10.4 Outlook 256
- 10.5 Conclusions 258
- References 258

11 Application of Engineered Biocatalysts for the Synthesis of Active Pharmaceutical Ingredients (APIs) 265

Juan Mangas-Sanchez, Sebastian C. Cosgrove, and Nicholas J. Turner

- 11.1 Introduction 265
 - 11.1.1 Transferases 266
 - 11.1.1.1 Transaminases 266
 - 11.1.1.2 Oxidoreductases 267
 - 11.1.1.2.1 Ketoreductases 267
 - 11.1.1.2.2 Amino Acid Dehydrogenases 271
 - 11.1.1.2.3 Cytochrome P450 Monooxygenases 272
 - 11.1.1.2.4 Baeyer–Villiger Monooxygenases 273
 - 11.1.1.2.5 Amine Oxidases 274
 - 11.1.1.2.6 Hydroxylases 276
 - 11.1.1.2.7 Imine Reductases 276
 - 11.1.1.3 Lyases 278
 - 11.1.1.3.1 Ammonia Lyases 278
 - 11.1.1.4 Isomerases 278
 - 11.1.1.5 Hydrolases 279
 - 11.1.1.5.1 Esterases 279
 - 11.1.1.5.2 Haloalkane Dehalogenase 279
 - 11.1.1.6 Multi-enzyme Cascade 281
 - 11.2 Conclusions 282
 - References 287

12	Directing Evolution of the Fungal Ligninolytic Secretome	295
	<i>Javier Viña-Gonzalez and Miguel Alcalde</i>	
12.1	The Fungal Ligninolytic Secretome	295
12.2	Functional Expression in Yeast	297
12.2.1	The Evolution of Signal Peptides	297
12.2.2	Secretion Mutations in Mature Protein	300
12.2.3	The Importance of Codon Usage	301
12.3	Yeast as a Tool-Box in the Generation of DNA Diversity	302
12.4	Bringing Together Evolutionary Strategies and Computational Tools	305
12.5	High-Throughput Screening (HTS) Assays for Ligninase Evolution	306
12.6	Conclusions and Outlook	309
	Acknowledgments	309
	References	310
13	Engineering Antibody-Based Therapeutics: Progress and Opportunities	317
	<i>Annalee W. Nguyen and Jennifer A. Maynard</i>	
13.1	Introduction	317
13.2	Antibody Formats	318
13.2.1	Human IgG1 Structure	318
13.2.2	Antibody-Drug Conjugates	319
13.2.3	Bispecific Antibodies	320
13.2.4	Single Domain Antibodies	321
13.2.5	Chimeric Antigen Receptors	321
13.3	Antibody Discovery	322
13.3.1	Antibody Target Identification	322
13.3.1.1	Cancer and Autoimmune Disease Targets	323
13.3.1.2	Infectious Disease Targets	323
13.3.2	Screening for Target-Binding Antibodies	324
13.3.2.1	Synthetic Library Derived Antibodies	324
13.3.2.2	Host-Derived Antibodies	325
13.3.2.3	Immunization	325
13.3.2.4	Pairing the Light and Heavy Variable Regions	326
13.3.2.5	Humanization	327
13.3.2.6	Hybrid Approaches to Antibody Discovery	328
13.4	Therapeutic Optimization of Antibodies	328
13.4.1	Serum Half-Life	328
13.4.1.1	Antibody Half-Life Extension	329
13.4.1.2	Antibody Half-Life Reduction	331
13.4.1.3	Effect of Half-Life Modification on Effector Functions	331

13.4.2	Effector Functions	331
13.4.2.1	Effector Function Considerations for Cancer Therapeutics	332
13.4.2.2	Effector Function Considerations for Infectious Disease Prophylaxis and Therapy	333
13.4.2.3	Effector Function Considerations for Treating Autoimmune Disease	334
13.4.2.4	Approaches to Engineering the Effector Functions of the IgG1 Fc	334
13.4.3	Tissue Localization	335
13.4.4	Immunogenicity	335
13.4.4.1	Reducing T-Cell Recognition	336
13.4.4.2	Reducing Aggregation	336
13.5	Manufacturability of Antibodies	336
13.5.1	Increasing Antibody Yield	337
13.5.1.1	Codon Usage	337
13.5.1.2	Signal Peptide Optimization	337
13.5.1.3	Expression Optimization	338
13.5.2	Alternative Production Methods	338
13.6	Conclusions	339
	Acknowledgments	339
	References	339

14 Programming Novel Cancer Therapeutics: Design Principles for Chimeric Antigen Receptors 353

Andrew J. Hou and Yvonne Y. Chen

14.1	Introduction	353
14.2	Metrics to Evaluate CAR-T Cell Function	354
14.3	Antigen-Recognition Domain	356
14.3.1	Tuning the Antigen-Recognition Domain to Manage Toxicity	356
14.3.2	Incorporation of Multiple Antigen-Recognition Domains to Engineer “Smarter” CARs	356
14.3.3	Novel Antigen-Recognition Domains to Enhance CAR Modularity	359
14.3.4	Engineering CARs that Target Soluble Factors	360
14.4	Extracellular Spacer	360
14.5	Transmembrane Domain	362
14.6	Signaling Domain	362
14.6.1	First- and Second-Generation CARs	362
14.6.2	Combinatorial Co-stimulation	363
14.6.3	Other Co-stimulatory Domains: ICOS, OX40, TLR2	364
14.6.4	Additional Considerations for CAR Signaling Domains	364
14.7	High-Throughput CAR Engineering	366
14.8	Novel Receptor Modalities	367
	References	369

Part IV Applications in Medical Biotechnology 377**15 Development of Novel Cellular Imaging Tools Using Protein Engineering 379***Praopim Limsakul, Chi-Wei Man, Qin Peng, Shaoying Lu, and Yingxiao Wang*

- 15.1 Introduction 379
- 15.2 Cellular Imaging Tools Developed by Protein Engineering 380
 - 15.2.1 Fluorescent Proteins 380
 - 15.2.1.1 The FP Color Palette 380
 - 15.2.1.2 Photocontrollable Fluorescent Proteins 381
 - 15.2.1.3 Other Engineered Fluorescent Proteins 383
 - 15.2.2 Antibodies and Protein Scaffolds 383
 - 15.2.2.1 Antibodies 383
 - 15.2.2.2 Antibody-Like Protein Scaffolds 384
 - 15.2.2.3 Directed Evolution 384
 - 15.2.3 Genetically Encoded Non-fluorescent Protein Tags 385
- 15.3 Application in Cellular Imaging 386
 - 15.3.1 Cell Biology Applications 386
 - 15.3.1.1 Localization 386
 - 15.3.1.2 Cell Signaling 387
 - 15.3.2 Application in Diagnostics and Medicine 390
 - 15.3.2.1 Detection 390
 - 15.3.2.2 Screening for Drugs 392
- 15.4 Conclusion and Perspectives 393
- References 394

Index 403

Part I

Directed Evolution

1

Continuous Evolution of Proteins *In Vivo*

Alon Wellner¹, Arjun Ravikumar¹, and Chang C. Liu^{1,2,3}

¹University of California, Department of Biomedical Engineering, 3201 Natural Sciences II, Irvine, CA, 92697, USA

²University of California, Department of Chemistry, 1102 Natural Sciences 2, Irvine, CA, 92697, USA

³University of California, Department of Molecular Biology and Biochemistry, 3205 McGaugh Hall, Irvine, CA, 92697, USA

1.1 Introduction

Directed evolution is a powerful approach for engineering new biomolecular and cellular functions [1–3]. In contrast to rational design approaches, directed evolution exploits diversity and evolution to shape the behavior of biological matter by applying the Darwinian cycle of mutation, selection, and amplification of genes and genomes. By doing so, the field of directed evolution has generated important insights into the evolutionary process [4–6] as well as useful RNAs, proteins, and systems with wide-ranging applications across biotechnology and medicine [7–11].

To mimic the evolutionary process, classical directed evolution approaches carry out cycles of *ex vivo* diversification on genes of interest (GOIs), transformation of the resulting gene libraries into cells, and selection of the desired function (Figure 1.1). Each iteration of this cycle is defined as a round of evolution, and as selection stringency increases over rounds, either automatically through competition or manually through changing conditions (or both), this process can lead GOIs closer and closer to the desired function. This overall process makes practical sense for a number of reasons, especially for the goal of protein engineering (i.e. GOI encodes a protein). First, *ex vivo* diversification is appropriate, because test tube molecular biology techniques such as DNA shuffling, site-directed saturation mutagenesis, and error-prone (ep) polymerase chain reaction (PCR) [2] are capable of generating exceptionally high and precise levels of sequence diversity for any GOI. Second, transforming diversified libraries of the GOI into cells is appropriate, because each GOI variant needs to be translated into a protein in order to express its function, and cells, especially model microbes, are naturally robust hosts for protein expression. Third, carrying out selection inside cells is appropriate, because (i) cells automatically maintain the genotype–phenotype connection between the GOI and expressed protein that is necessary for amplification of desired variants,

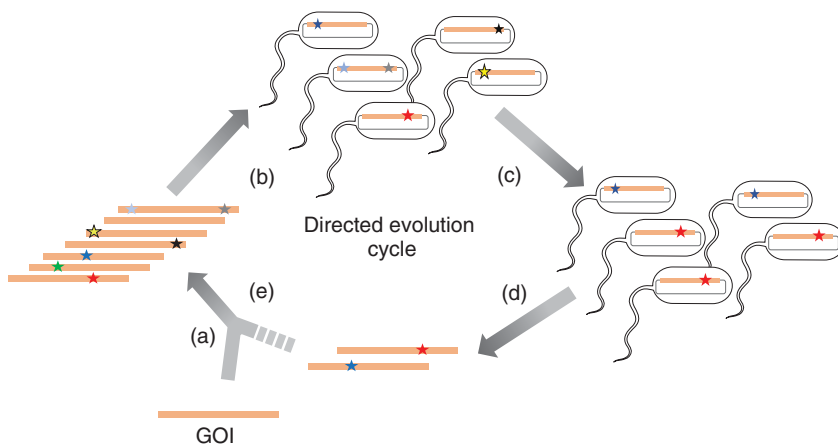


Figure 1.1 A schematic illustration of a typical directed evolution setup. (a) A GOI is diversified *ex vivo*, typically by applying an error-prone PCR to generate a GOI library. (b) The library is then cloned into an expression vector and transformed/transfected into cells that are subjected to (c) outgrowth and selection for enhanced protein activity. (d) Plasmid DNA that is enriched for library members with increased properties is extracted and (e) subjected again to diversification and selection. The directed evolution cycle is iterated until the desired outcome is achieved or until diminishing returns (a plateau is reached).

(ii) we often care about a GOI's function within the context of a cell, especially as metabolic engineering and cell-based therapy applications mature, and (iii) the use of cell survival as the output for a desired protein function allows millions or billions of GOI variants to be simultaneously tested – it is easy to culture billions of cells under selection conditions – in contrast to *ex vivo* screens that are much lower throughput. Survival-based selections are not always immediately available, but one can often find a way to reliably link the desired function of a protein to cellular fitness.

While sensible, the practical requirement that diversification should occur *in vitro* but expression and selection should occur *in vivo* in this classical directed evolution pipeline creates significant suboptimalities. First, the number of steps that can be taken along an adaptive path becomes few, since each round of *in vitro* mutation, transformation, and *in vivo* selection takes several days or weeks to carry out. Second, limited DNA transformation efficiencies result in strong bottlenecks of diversity that can mitigate the probability of finding the most optimal solutions in sequence space. Third, the number of evolution experiments that can be run simultaneously is minimal, because *in vitro* mutagenesis, cloning, and transformation are experimentally onerous, demanding extensive researcher intervention [12]. These shortcomings keep two highly promising categories of experiments largely outside the grasp of classical methods: first is the directed evolution of genes towards highly novel functions that likely require long mutational paths to reach (e.g. the optimization of multi-gene metabolic pathways or the *de novo* evolution of enzyme activity); and second is the large-scale replication of directed evolution experiments, needed in cases when many different functional variants of a gene are desired (e.g. the evolution of

multiple synthetic receptors for a collection of ligands) or when statistical power is required in order to understand outcomes in experimental evolution (e.g. probing the scope of adaptive trajectories leading to resistance in a drug target).

An emerging field of *in vivo* continuous directed evolution seeks to overcome these shortcomings by performing both continuous diversification of the GOI and selection entirely within living cells [13]. In this way, GOIs can be rapidly and continuously evolved through basic serial passaging of cells under selective conditions. This removes the labor-intensive cycling between *in vitro* and *in vivo* steps and the DNA transformation bottlenecks associated with the classical pipeline, creating a new paradigm for directed evolution that is limited only by the generation time of the host cell and the number of cells that can be cultured. These limitations are usually negligible – in most host organisms for directed evolution such as *Escherichia coli* and *Saccharomyces cerevisiae*, generation time is fast (20–100 minutes) and the number of cells that can be cultured is massive (10^8 – 10^9 ml⁻¹) – so the potential power of continuous systems is enormous. Moreover, *in vivo* continuous directed evolution is amenable to high-throughput experiments, because serial passaging is straightforward and can be automated at scale or converted to continuous culture using bioreactors [14–16]. In this chapter, we discuss various systems that partially or fully achieve *in vivo* continuous evolution (ICE).

1.2 Challenges in Achieving *In Vivo* Continuous Evolution

Before discussing how ICE can be realized, we shall first clarify why this is a challenging problem. The difficulty of achieving ICE of GOIs lies in the fundamental relationship between how fast one can mutate an information polymer and its length. Several theories predict that organisms face an “error threshold” at mutation rates on order $1/L$ (where L is the length of the genome), near which selection cannot maintain fitness, leading to gradual decline towards low fitness, or above which one is nearly guaranteed a lethal mutation every cycle of replication, leading to rapid extinction [17–20]. Because cellular genomes are large (e.g. $\sim 5 \times 10^6$ in *E. coli*, $\sim 1.2 \times 10^7$ in *S. cerevisiae*, and $\sim 3 \times 10^9$ in humans), this implies that evolution strongly favors low genomic mutation rates (e.g. $\sim 5 \times 10^{-8}$ substitutions per base [s.p.b.] in *E. coli*, $\sim 10^{-10}$ s.p.b. in *S. cerevisiae* and $\sim 3 \times 10^{-9}$ in human somatic cells) [21–23]. Experiment confirms this prediction. Drake observed empirically that mutation rates scale as $1/L$ across many organisms [17]; evolution experiments have shown that when mutator phenotypes do arise, they are accompanied by fitness costs and only transiently persist [19, 24–26]; and more direct tests in yeast find that there is indeed a mutation-induced extinction threshold at $\sim 1/L$, above which yeast cannot propagate [18]. Yet individual GOIs are small in comparison with genomes, so they are capable of tolerating much higher error rates. In fact, they *require* much higher per base error rates than genomes to generate the same amount of total mutational diversity, because they have fewer bases. Following the $1/L$ scaling, a typical 1 kb GOI should be able to tolerate mutation rates on order $\sim 10^{-3}$ s.p.b.

Therefore, the primary challenge in achieving rapid ICE is how to develop molecular machinery or other strategies that target rapid mutagenesis to only GOIs, allowing the host genome to replicate at mutation rates below its low error thresholds but driving the GOI at the high mutation necessary for fast generation of sequence diversity. When considering the level of targeting in the ideal case, the formidability of this challenge becomes quite apparent. Ideally, one should continuously mutate GOIs at rates close to their error threshold ($\sim 10^{-3}$) to maximize diversification but leave the genomic error rate completely unchanged, as the genome's error rate is evolutionarily optimized for host fitness. In *E. coli*, *S. cerevisiae*, and human cells, this means that on-target versus off-target mutagenesis must differ by 10^6 -fold, 10^7 -fold, and 10^7 -fold, respectively, which is much more than the 10- to 1000-fold targeting required in most synthetic biology problems involving molecular recognition. *How can we achieve such extreme precision in mutational targeting in the cell?*

There is yet another hard challenge in realizing ICE, which has to do with the durability of mutagenesis. Ideally, one wants a high rate of mutagenesis on the GOI to persist indefinitely (or at least for as long as the experimenter cares), so that a protein can traverse long mutational pathways towards desired functions. Because one needs to achieve mutational targeting to the GOI, there is almost always a risk to durability: any mechanism for targeting the GOI over the rest of the genome will necessarily rely on some cis-elements in or surrounding the GOI to mediate the targeting. If these cis-elements become mutated, which is quite likely since they are usually in or near the GOI undergoing rapid mutation, then mutagenesis will slow or stop. Ideally, a continuous evolution system will limit the chance that a cis-element for mutational targeting gets degraded. In the case that it does, an ideal system will remove the GOI containing the mutated cis-element from the population so that it can't fix in the population (through gradual mutational accumulation or a selective sweep if mutagenesis comes with a fitness cost) and end the continuous evolution process prematurely. *How do we achieve architectures for durability?*

Other challenges for ICE include generality across host organisms, the ability to mutate many genes simultaneously, and fine control over mutation rate and spectra; but the most defining ones are targeting and durability. In the remainder of this chapter, we review several *in vivo* continuous directed evolution platforms within the framework of these challenges. We highlight in Section 1.4.4 and note here that our recently developed orthogonal DNA replication (OrthoRep), among systems for ICE, seems uniquely capable of complete precision in mutational targeting (as far as we can tell), and is a highly durable architecture for enforcing prolonged mutagenesis in GOIs. We also highlight, in Section 1.3, phage-assisted continuous evolution (PACE), which has been remarkably successful for continuous biomolecular evolution. Although PACE is not an entirely *in vivo* system, it achieves complete precision in mutational targeting and durability – in fact by not being entirely *in vivo*, as we will explain. We do not discuss several powerful technologies for non-continuous *in vivo* diversification or streamlined diversification methods, such as MAGE [27], CREATE [28], DIVERGE [29], and CPR [30], but note that these are also promising approaches to protein evolution as they address some of the constraints of classical directed evolution methods. A summary of various characteristics of the systems we discuss is provided in Table 1.1.

Table 1.1 Comparison among approaches for *in vivo* continuous evolution.

Approach	Systems	Mutation rate	Targeting of mutagenesis	Durability of mutagenesis	Number (simple estimates) and location of genes that can be evolved simultaneously	Generality across host organisms	Mutational spectrum	References
Continuous rounds of evolution with a conditionally replicating bacteriophage	PACE	Mutates GOIs at $\sim 10^{-3}$ s.p.b.	Complete targeting to the bacteriophage genome, since <i>E. coli</i> are constantly replaced	Indefinitely continuous since mutagenesis is enforced. In practice, this method is typically implemented for 1–3 weeks	1–10 genes encoded on bacteriophage genome.	Currently in <i>E. coli</i> . Could be implemented with mammalian cells using non-integrating viruses (e.g. adenovirus).	Fairly unbiased mutational spectrum.	[33]
Targeted mutagenesis in <i>E. coli</i> with error-prone DNA polymerase I	ep Pol I/ColE1-based systems, CRISPR-guided DNA polymerases (EvolvR)	GOIs encoded near the ColE1 origin are mutated by ep Pol I at $\sim 10^{-3}$ s.p.b. CRISPR-guided Pol I can induce rates as high as 10^{-2} , but this quickly drops off after the guide region	Targeting with unfused ep Pol I is maximally only ~ 400 -fold. Fusion to nCas9 generally improves targeting to ~ 1000 -fold	Durability remains to be tested. Ep Pol I/ColE1 incurs significant off-target mutagenesis, which could quickly abrogate mutagenesis. EvolvR risks breaking down because it rapidly mutates the gRNA target region	1–5 genes encoded on a plasmid with ep Pol I/ColE1. 1–20 genes on plasmids or at their endogenous genomic loci with EvolvR, depending on how many targeting sgRNAs one can stably encode.	Both systems are currently in <i>E. coli</i> . EvolvR should be fairly general across hosts, especially with the use of Phi29 DNAP.	ep Pol I mutates ColE1 plasmids with a bias towards transition mutations. EvolvR generates substitutions of all four nucleotide types, in a relatively unbiased manner. If needed, this can be improved through DNAP engineering.	[47–50, 54]

(Continued)

Table 1.1 (Continued)

Approach	Systems	Mutation rate	Targeting of mutagenesis	Durability of mutagenesis	Number (simple estimates) and location of genes that can be evolved simultaneously	Generality across host organisms	Mutational spectrum	References
Yeast systems that do not use engineered DNA polymerases for mutagenesis	TaGTEAM, ICE	For TaGTEAM, $\sim 10^{-4}$ s.b.p. at 10 kb regions on both sides of the tetO array. For ICE, 1.5×10^{-4} , if excluding the rate of retrotransposition needed to induce mutagenesis	TaGTEAM offers targeting of genomic GOIs, however with low accuracy. ICE's targeting is theoretically good since off-target regions are not reverse transcribed	Durability remains to be tested. Off-target mutation and the requirement that retrotransposition occurs back into the original locus for continued evolution with ICE will likely affect durability	1–10 genes on plasmids or at engineered genomic loci	Both systems are currently in yeast. ICE has been demonstrated in several diverged yeast species. TaGTEAM should function in <i>E. coli</i> and mammalian cells. ICE could be implemented in new hosts using retrotransposable elements similar to Ty1.	TaGTEAM generates a broad spectrum of both transitions and transversions. In addition, 25% of mutations are single base deletions. In ICE there is a 1 : 1 ratio between transitions and transversions.	[55, 58]

Somatic hypermutation as a means for targeted mutagenesis of GOIs	Hypermutator B cell line, Ramos cell line, dCas9-AID fusions (such as CRISPRx), T7 RNAP-AID fusion	CRISPRx mutates GOIs at $\sim 5 \times 10^{-4}$ s.p.b.	Efficient targeting. No increase in mutagenesis rate was detected in an off-target locus. The hyperactive AID variant can create dense, highly variable point mutations within a region of 100 bp surrounding an sgRNA target site	Durability remains to be tested.	1–10 genes on plasmids or at engineered genomic loci with the hypermutator B cell line, Ramos cell line, or T7 RNAP-AID fusion. Dozens of genes at endogenous genomic loci with dCas9-AID fusions	Systems depending on natural SHM are limited to mammalian cells. AID-fusions are currently available in mammalian systems or <i>E. coli</i> , depending on the system. AID fusions should be extensible to all host-types.	AID generates point mutations rather than insertions and deletions, and it favors transitions over transversions. However, repair pathways operate at AID-mutated loci to extend the scope of mutagenesis.	[67–73]
Orthogonal DNA replication	OrthoRep	Mutates GOIs at $\sim 10^{-5}$ s.p.b.	Complete orthogonality (at least 100 000-fold targeting)	Indefinitely continuous since mutagenesis is enforced. This method has been implemented for up to 300 generations without any sign of erosion	1–10 genes encoded on a special orthogonal plasmid	Currently in yeast. Should be extensible to bacteria and mammalian systems using related protein-priming DNAPs.	TP-DNAP1-4-2 strongly favors transition mutations. This can be readily improved through DNAP engineering.	[74, 75, 78]

Source: Esvelt et al. [33]; Fabret et al. [47]; Alexander et al. [49].

1.3 Phage-Assisted Continuous Evolution (PACE)

The most successful method for continuous protein evolution thus far is the PACE system developed in the lab of David Liu (Figure 1.2) [2, 12, 14, 15, 31–37]. PACE reimagines traditional “rounds” of directed evolution as generations of the M13 bacteriophage life-cycle, thereby transforming a step-wise and labor-intensive procedure into a continuous biological process. In PACE, GOIs are encoded in the M13 genome, and the resulting phage continuously replicate in a vessel (termed “lagoon”) that experiences a constant influx of *E. coli* cells. To create a selection pressure for GOIs to evolve, the activity of interest is coupled to phage survival. This is achieved by deleting the essential gene III (gIII), encoding coat protein III (pIII), from the M13 genome. The host *E. coli* strain is engineered to encode gIII in a genetic circuit that makes pIII expression dose-dependent on the desired activity of the GOI (see the following text for examples); so only phage that successfully evolves the GOI can trigger pIII expression and continue propagating. Due to the rapid generation time of M13 (~10 minutes without selection), evolution in this manner can iterate hundreds of times in just a few days.

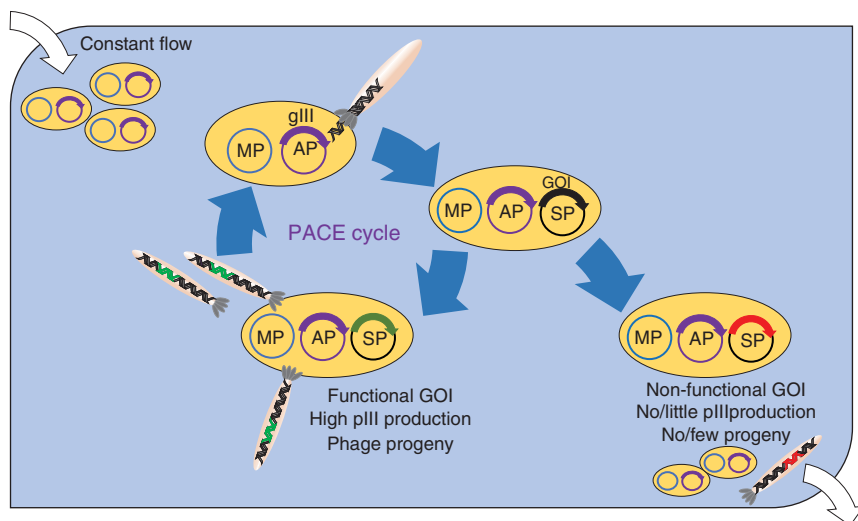


Figure 1.2 Pace. Phage carrying the selection plasmid (SP) encoding the GOI propagates on *E. coli* cells which are constantly flowing into the “lagoon” at a rate that does not permit their propagation but is longer than the phage life cycle, thus permitting phage replication. Upon infection, the SP (as well as the bacterial genome) experiences a high degree of mutagenesis due to the presence of a mutator plasmid (MP). In a PACE experiment, high GOI activity (green) is linked to drive strong gIII expression, resulting in progeny that can then infect incoming *E. coli*. No GOI activity (or a weak one, red) results in poor progeny production, becoming washed away from the lagoon at a larger rate (alongside bacterial cells). The system is designed to run for hundreds of generations without human intervention and result in the evolution of the GOI towards the desired activity. Source: Packer and Liu [2]; Badran and Liu [12]; Carlson et al. [14]; Dickinson et al. [15].

A key parameter in PACE is the *E. coli* flowrate, which should exceed their doubling time but be slower than the phage life cycle, allowing only phage to replicate in the lagoon (on average). Consequently, only phage accumulates mutations, whereas *E. coli* are physically prevented from doing so. High rates of mutation on the phage (and *E. coli*) genome are driven by a mutator plasmid (MP) that is carried by the *E. coli* cells and induced in the lagoon for error-prone M13 replication. The latest version of the MP is able to drive potent mutagenesis at $>10^{-3}$ s.p.b. by combining the effects of six different mutagenesis drivers [38].

Esvelt et al. first demonstrated proof of concept by evolving T7 RNA polymerase (RNAP) to initiate transcription from new promoter sequences [33]. pIII expression was bottlenecked at the level of transcription by encoding promoter sequences unrecognized by wild-type (wt) T7 RNAP (or any *E. coli* RNAPs), thus driving the selection to favor T7 RNAP variants that are able to efficiently recognize the new promoters. After eight days and 200 “rounds” of PACE, new T7 RNAPs emerged that could transcribe from the distant T3 RNAP promoter as efficiently as wt T7 RNAP does from its cognate promoter [33]. Similarly, T7 RNAP variants that efficiently initiate transcription with ATP or CTP, instead of GTP, were evolved. Since that landmark study, the ability to couple T7 RNAP activity to PACE has been exploited in a number of ways, ranging from basic adaptation studies to selections for split T7 RNAP [14, 15, 35–38].

In principle, PACE is applicable for the evolution of any biomolecular function that can be linked to pIII expression; and in just a few years since its inception, this has been realized in a wide range of applications beyond RNAP evolution. A notable example is the evolution of new DNA binding domains. Hubbard et al. employed the classic one-hybrid selection with PACE to evolve transcription activator-like effector nucleases (TALENs) with broadly improved DNA cleavage specificity [34]. Although TALENs are highly promising for gene editing, their major limitation is that they require the 5′ nucleotide of the target sequence to be T [39]. New TALEs (TALENs without the fused nuclease) were evolved with PACE by fusing the DNA binding domain of the canonical CBX8-targeting TALE to the ω subunit of *E. coli* RNAP. The PACE system was designed to include the TALE target sequence upstream of gIII. TALEs that successfully bind the target DNA recruit holoenzyme RNAP around the ω subunit, resulting in subsequent pIII expression. With this TALE selection, the identity of the target sequence can be custom-tailored, in this case, to encode noncanonical 5′ nucleotides. After using an additional negative selection (see below) that inhibited variants with promiscuous substrate specificity, Hubbard et al. were able to evolve TALE variants that displayed two- to fourfold increases in specificity for 5′ A, 5′ C, or 5′ G versus 5′ T, relative to wt TALE.

The one-hybrid PACE format was also used for overcoming one of Cas9’s main limitations, restricted protospacer adjacent motif (PAM) compatibility. This time, Hu et al. fused a catalytically dead variant of *Streptococcus pyogenes* Cas9 (dCas9) to the ω subunit of *E. coli* RNAP [40]. Then, the authors cleverly fed the lagoon with a mixture of host *E. coli* cells bearing a library of target sequences that covers all 64 possible PAM sequences, to select for broadened PAM compatibility. After PACE, several variants were isolated that could efficiently recognize NG, GAA, and

GAT as PAMs. Upon restoration of nuclease catalytic activity to these evolved dCas9 variants, the authors remarkably found that one of them, xCas9, exhibited greater DNA specificity than wt Cas9, even with its newly-gained broad PAM compatibility. This result challenges the widely-held assumption that there must be a trade-off between editing specificity and PAM compatibility and suggests that Cas9 can be improved through laboratory evolution to meet the most demanding challenges of CRISPR-Cas9 applications.

Another important form of PACE is its use with two-hybrid selection for the evolution of high-affinity protein-binders [31]. In the bacterial two-hybrid system, the ω subunit of *E. coli* RNAP is fused to a protein of interest, which is recruited to DNA through its interaction with a target protein. This target protein is fused to a DNA binding domain that localizes the complex at its cognate sequence encoded upstream of a reporter gene. If the protein of interest binds the target protein, then the RNAP holoenzyme can reconstitute around the ω subunit and drive expression of the downstream reporter. Badran et al. adapted this system for PACE using gIII as the reporter. After extensive optimization, Badran et al. were able to use this PACE format to evolve the insecticidal protein, *Bacillus thuringiensis* δ -endotoxin (Bt toxin) Cry1Ac, to bind and inhibit a new receptor in the gut of the insect pest *Trichoplusia ni* (TnCAD) [31]. Although wt Cry1Ac did not detectably bind TnCAD, the evolved variants were able to bind with nM affinity. Significantly, this strategy could overcome widespread Bt toxin resistance, which primarily occurs through mutational changes that inhibit binding to the native receptor of wt Cry1Ac. Badran et al. demonstrated this by showing that evolved Cry1Ac is highly potent at killing *T. ni* that are resistant to wt Cry1Ac. An exciting possibility for the future would be to evolve TnCAD to resist the new Cry1Ac variant, and then iterate this cycle in a study of molecular co-evolution.

Additional positive selections developed for PACE have enabled evolution of proteases that are drug resistant [32] or have altered substrate specificities [41], aminoacyl-tRNA synthetases (aaRSs) that can accept noncanonical amino acids [42], and protein variants with improved soluble expression [43]. Negative selections are also compatible with PACE, and are useful in cases where it is desirable to evolve high specificity towards the target substrate and restrict promiscuity towards others (especially the native substrate). This can be achieved by introducing a dominant negative allele of pIII, pIII-neg, that inhibits phage propagation [14]. The expression of pIII-neg can then be linked to the unwanted activity (e.g. recognition of the T7 promoter by T7 RNAP) for negative selection. (This strategy was successfully employed during TALEN and aaRS evolution.) Selection stringency and mutation rate are also important determinants of PACE outcomes and can be titrated [14, 35]. Lastly, we note that the Isalan lab developed a system related to PACE that accommodates the evolution of multiple genes, starting from combinatorial libraries. With this system, they were able to evolve a panel of orthogonal dual promoter-transcription factor pairs that were used to make multi-input logic gates [44, 45].

Clearly, PACE is a powerful method for continuous protein evolution, but as noted early in this chapter, it is not an entirely *in vivo* system. Rather, M13 serves as a

biological carrier of the GOI from one *E. coli* host cell to the next, with a given cell serving as a host of error-prone replication just once (on average). This ingenious design circumvents the challenges of *in vivo* mutational targeting. Since mutagenesis is induced in the lagoon, where *E. coli* briefly reside without doubling, mutation rates can be elevated entirely through untargeted mechanisms (and temporarily induced to be as high as desired), without consideration for replication of the *E. coli* genome. Even if *E. coli* cells stochastically replicate in the lagoon and become a source of cheater mutations (e.g. constitutive gIII expression), the flow rate ensures that any progeny are quickly diluted out. What's left in the lagoon is a population of M13 that selectively undergoes error-prone replication. In effect, targeting of mutations to the phage genome containing the GOI is complete, as the host *E. coli* is constantly replaced.

PACE also achieves durable mutagenesis by enforcing continuity. Replication of GOIs is intrinsically coupled to mutagenesis, through error-prone replication of the M13 genome. Any phage that escapes mutagenesis through a mutation in the phage genome's origin of replication, for example, must do so at the expense of being replicated. Only variants that continue to accumulate mutations can survive and propagate. And since *E. coli* cells do not persist long enough in the lagoon to evolve, the mutation rate experienced by phage remains unchanged. The durability of PACE is best evidenced by the long mutational trajectories traversed during evolution experiments, which have yielded protein variants with up to 16 mutations [46].

However, because PACE is not entirely *in vivo*, it suffers two major limitations. First, it requires continuous propagation of phage in a population of freshly diluted *E. coli* cells, which has been achieved thus far with a chemostat or turbidostat setup. This greatly limits the throughput and accessibility of PACE experiments, typically to fewer than ten replicates or experiments especially when different selection environments are desired across replicates. Second, PACE is restricted to selections that are linked to phage propagation. This precludes selections for *in vivo* phenotypes like tolerance or metabolism, as well as cell-based selections like fluorescence-activated cell sorting (FACS) or droplet sorting. These limitations motivate the need for continuous directed evolution systems that operate entirely *in vivo*.

1.4 Systems That Allow *In Vivo* Continuous Directed Evolution

1.4.1 Targeted Mutagenesis in *E. coli* with Error-Prone DNA Polymerase I

The first system that was able to perform continuous targeted mutagenesis *in vivo* was published in 2000 by Fabret et al. [47]. It was designed based on the developments in understanding the mechanism of ColE1 plasmid replication in *E. coli*. For plasmids that contain a ColE1 origin of replication, DNA polymerase (DNAP) I (Pol I) is responsible for elongating from the RNA primer that initiates replication at the origin. Pol I will extend for about 400–2000 bp, after which DNAP III

(Pol III), responsible for bulk DNA replication in *E. coli*, replaces Pol I [48]. When using a genome-encoded proofreading-deficient Pol I, genes that were cloned near the ColE1 origin experienced a 6- to 20-fold higher degree of mutagenesis over genes at more remote areas in the plasmid, showing targeting. The system's components were further combined with mismatch repair mutants to raise the mutation rate on GOIs yet another 20- to 40-fold, although significant increases in genomic mutation rates of at least several hundred-fold were observed. As a proof of concept, the authors evolved dominant negative variants of LacI that would outcompete a genomically-encoded wt LacI in binding its cognate operator, LacO. After 30 generations, LacI mutants that caused complete abolishment of wt LacI's binding to LacO were isolated. These variants were altered in their DNA binding domain but still formed tetramers with wt LacI, thereby abolishing LacI's repression at LacO.

Further improvement of the Pol I/ColE1 system was demonstrated in 2003 (Figure 1.3a) [46, 49]. Camps et al. modified the system to express the ep Pol I from a plasmid with a Pol I-independent origin of replication. Then, they used a host *E. coli* strain (J2000) whose genomically-encoded wt Pol I was temperature sensitive (ts) [49]. At restrictive temperatures, the ts Pol I becomes inactive such that only the ep Pol I acts, preventing the high-fidelity ts Pol I from competing for replication at the ColE1 origin. Based on prior studies of Pol I from the same lab [50], Camps et al. engineered a Pol I variant that was exceptionally error-prone, leading to mutation rates as high as 8.1×10^{-4} s.p.b at the GOI when the ts Pol I was inactivated. Mutagenesis expanded to about 3 kb from the ColE1 origin and

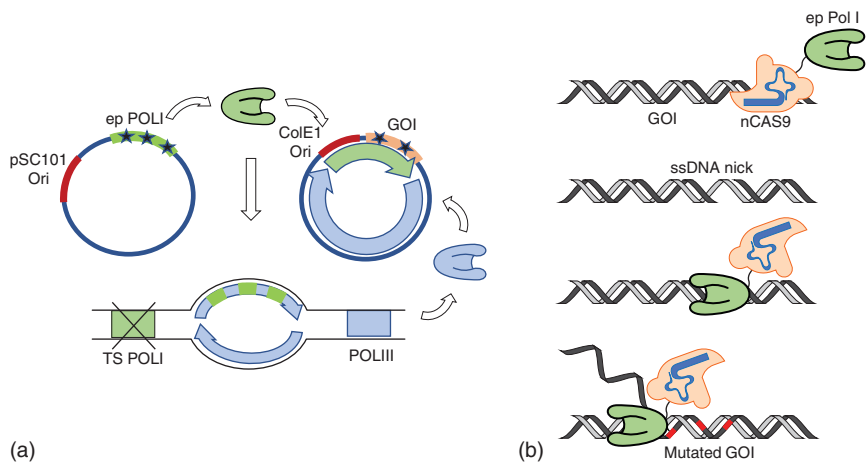


Figure 1.3 Targeted mutagenesis in *E. coli* with error-prone DNA polymerase I. (a) An ep version of Pol I is expressed from a plasmid whose replication is driven by a non-ColE1 origin of replication (ori). The GOI is placed on the target plasmid near the ColE1 ori and thus targeted for mutagenesis. After 1–3 kb of ep replication, Pol III replaces Pol I to replicate the remainder of the plasmid with high fidelity. The genomic allele of POL I is temperature sensitive, such that enhanced mutagenesis can be induced by growth at the restrictive temperature. Source: Alexander et al. [49]; Camps et al. [46]. (b) The EvolvR system is composed of a CRISPR-guided nickase that nicks the target GOI, fused to ep Pol I that performs nick translation.

was evenly distributed within this region, albeit with certain biases in mutational preference. As a proof of concept experiment, Camps et al. demonstrated that their system could be used to evolve enzymes with diverged function by generating TEM-1 β -lactamase mutants that were able to hydrolyze a third-generation lactam antibiotic, aztreonam.

The ep Pol I/ColE1 system has subsequently been applied in a handful of additional directed evolution experiments. For example, Koch et al. used the system to prepare a library of terminal alkane hydroxylases with the aim of evolving variants that can oxidize butane [51]. Although they only used the system for the preparation of mutant libraries (i.e. as a mutator strain) and not for continuous evolution involving serial passaging under prolonged selection conditions, they demonstrated that one can create large libraries of GOI variants directly *in vivo*. In another application, an M13 phagemid with a ColE1 origin was made to encode LuxR and infect *E. coli* harboring the ep Pol I [52]. LuxR is a transcriptional activator and drove the transcription of an antibiotic resistance gene (β -lactamase) controlled by the *lux* promoter in the *E. coli*. Through several cycles of infecting fresh *E. coli*, antibiotic selection, lysis of *E. coli*, and phage isolation, LuxR evolved a 17-fold higher binding affinity to the *lux* promoter sequence.

While the ep Pol I/ColE1 system approaches ICE, it is limited by off-target mutagenesis and low durability. Because Pol I is responsible for Okazaki fragment mending throughout the genome and also participates in DNA repair [53], expressing an ep Pol I causes substantial mutagenesis genome-wide. Targeting of mutations to the GOI does occur – owing to the ColE1 origin, the limited role of Pol I in lagging strand replication, and special growth conditions optimized to time ep Pol I action with growth phases where genome replication activity is low – but is maximally only \sim 400-fold. Therefore, when highly ep Pol Is are used, it is possible that off-target mutagenesis will lower the fitness of the cell, causing fixation of suppressor mutations that abrogate the activity of ep Pol I. Still, the Pol I/ColE1 system represents a landmark development that encouraged the field to pursue new strategies for realizing ICE.

Perhaps the closest conceptual descendant of the ep Pol I/ColE1 system is a new *E. coli* continuous evolution system called EvolvR, which uses CRISPR-guided ep DNAPs to continuously target mutations to GOIs (Figure 1.3b). Rather than rely on the natural targeting of Pol I to ColE1, Halperin et al. [54] fused ep Pol I variants (and other DNAPs) to a nickase Cas9 (nCas9) that would serve two purposes. First, nCas9 would bring the ep Pol I to any GOI encoded on a plasmid or the genome using a guide RNA (gRNA). Second, the nCas9 would nick the target strand, creating a free 3'-OH substrate from which the ep Pol I could extend. Once nCas9 releases the nicked product, it is believed that ep Pol I then latches on and carries out error-prone extension from the nick. This highly clever idea was demonstrated in *E. coli* with a number of ep Pol I variants spanning different mutation rates and activities, as well as with a moderately ep Phi29 DNAP with high processivity. Using the most mutagenic ep Pol I, Halperin et al. measured a mutation rate approaching 10^{-2} s.p.b. (a 7.7 million-fold elevation compared to wt cells) at the first nucleotide 3' of the nCas9-induced nick. While this extreme mutation rate quickly dropped when

moving away from the nick, other Pol I and Phi29 DNAP variants with moderate error rates could achieve mutagenesis windows up to 350 bp. With these characteristics and with the potential to use multiple gRNAs to simultaneously target multiple parts of a gene, EvolvR could readily and efficiently generate sequence diversity on a GOI *in vivo* to support continuous evolution. Indeed, in a proof of principle experiment, Halperin et al. used EvolvR to rapidly evolve spectinomycin resistance by targeting mutagenesis to the *rpsE* gene and found new resistance mutations that were previously unknown.

Future studies and improvements on EvolvR will clarify how well it drives ICE for prolonged periods of time, needed to traverse long mutational pathways. Durability may be difficult in the current architecture, because the mutation rate is maximal at nucleotides within the target region of the gRNA, which if mutated, will reduce the ability of the system to continue inducing mutagenesis. Since the GOI can still be replicated (by high-fidelity host systems) in the absence of EvolvR function, this may result in the fixation of partially adapted GOI mutants that stop mutating, leading to premature cessation of evolution. In addition, EvolvR still has off-target elevations in mutation rate, presumably because ep Pol I or Phi29 can participate in genomic replication and/or because Cas9 has off-target binding. Strategies that use more processive ep DNAPs with no activity in normal genome replication and alternative CRISPR systems that nick outside the critical regions for gRNA targeting may overcome potential issues of targeting and durability. We also anticipate that this system should readily transfer to cell-types other than *E. coli*. Therefore, EvolvR is a highly promising new system for ICE with enormous potential, especially for the multiplexed evolution of genes at their endogenous genomic loci rather than on a plasmid.

1.4.2 Yeast Systems That Do Not Use Engineered DNA Polymerases for Mutagenesis

The first demonstration of continuous targeted mutagenesis *in vivo* in yeast was published in 2013 under the name TaGTEAM (Figure 1.4a), which stands for targeting glycosylases to embedded arrays for mutagenesis [55]. In TaGTEAM, mutagenesis at the GOI is initiated by recruiting a DNA glycosylase, which normally functions as the first step in the base excision repair (BER) pathway responsible for removing chemically altered DNA bases [56]. The authors adopted the yeast 3-methyladenine glycosylase, Mag1p, and fused it to the tet repressor (tetR) that binds a 19-bp operator sequence, tetO. By introducing a non-recombinogenic tetO array (with each tetO site separated by 10–30 bp of random sequence), the tetR-Mag1p fusion could be targeted to GOIs in the chromosome or plasmid. It is presumed that tetR-Mag1p targeting generates a build-up of unprocessed abasic sites at target loci, leading to replication fork stalling and recruitment of ep translesion polymerases [57]. This faulty repair can lead to both point mutations and frameshifts. To test their system for its ability to generate mutagenesis at a GOI, Finney-Manchester et al. introduced a 240X tetO array upstream of a *URA3* auxotrophic marker in a region of chromosome 1 that does not contain nearby essential genes. The distance between the tetO array and the marker was titrated to assess the size of the area subjected to mutagenesis. The